

## ABSTRACT

A fast low voltage ballistic program, ultra-short channel, ultra-high density, dual-bit multi-level flash memory is described with a two or three polysilicon split gate side wall process. The structure and operation of this invention is enabled by a twin MONOS cell structure having an ultra-short control gate channel of less than 40nm, with ballistic injection which provides high electron injection efficiency and very fast program at low program voltages of 3~5V. The cell structure is realized by (i) placing side wall control gates over a composite of Oxide-Nitride-Oxide (ONO) on both sides of the word gate, and (ii) forming the control gates and bit diffusion by self-alignment and sharing the control gates and bit diffusions between memory cells for high density. Key elements used in this process are: 1) Disposable side wall process to fabricate the ultra short channel and the side wall control gate with or without a step structure, and 2) Self-aligned definition of the control gate over the storage nitride and the bit line diffusion, which also runs in the same direction as the control gate. The features of fast program, low voltage, ultra-high density, dual-bit, multi-level MONOS NVRAM of the present invention include: 1) Electron memory storage in nitride regions within an ONO layer underlying the control gates, 2) high density dual-bit cell in which there are two nitride memory storage elements per cell, 3) high density dual-bit cell can store multi-levels in each of the nitride regions, 4) low current program controlled by the word gate and control gate, 5) fast, low voltage program by ballistic injection utilizing the controllable ultra-short channel MONOS, and 6) side wall control poly gates to program and read multi-levels while masking out memory storage state effects of the unselected adjacent nitride regions and memory cells. The ballistic MONOS memory cell is arranged in the following array: each memory cell contains two nitride regions for one word gate, and  $\frac{1}{2}$  a source diffusion and  $\frac{1}{2}$  a bit diffusion. Control gates can be defined separately or shared together over the same diffusion. Diffusions are

- shared between cells and run in parallel to the side wall control gates, and perpendicular to the word line.